

Beszámoló az OTKA 68749 területi adatok elemzése kutatásáról

Napjainkban főként a térinformatika robbanásszerű terjedése miatt egyre gyakrabban alkalmazunk területi statisztikai módszereket. E módszerek többsége a bányászatban keletkezett, bányászati problémák megoldására. A szakterület gyors fejlődése azonban ma már sok más alkalmazási területet is felölel, így gyakorta találkozhatunk társadalomtudományi alkalmazásokkal is. Hazánkban a társadalomtudományi alkalmazások számának gyarapodása nem követte a külföldi gyakorlatot. Szakkönyvek sem jelennek meg.

Ez annál is sajnálatosabb, mivel megítélésem szerint a területi statisztikai módszerek alkalmazása a társadalomtudományokban és különösen szociológiában, ígéretes fejlesztési terület. E módszerek segítségével lehetségessé válhat eddig nem vagy kevéssé megoldott problémák kezelése.

Az aggregált adatokból az egyéni magatartásra történő következtetés vizsgálata majd száz évre nyúlik vissza. Durkheim Öngyilkosság c. könyvében már hasonló problémafelvetéssel él, hiszen különböző nemzetiségi, vallási adatokból kívánt az egyének magatartására következtetéseket levonni. Durkheim számos különböző franciaországi közigazgatási területre bontott térképet készített el és vizsgált meg részletesen. Különböző hipotéziseket elemzett. Az elmebetegség és az öngyilkosság kapcsolatát, az öngyilkosság és az alkoholizmus kapcsolatát vizsgálta. Durkheim rendelkezésére különböző területi egységre összevont statisztikai adatok álltak és ezekből az adatokból kívánt az egyes egyének öngyilkossági magatartására következtetéseket levonni, azaz mai szóhasználattal élve, az ökológiai kapcsolat problémájának úttörő elemzője volt.

Területi adatok jellemzői

A területi adatokat durván három csoportba sorolhatjuk: Az első csoportban a területileg folytonos adatokat vizsgáljuk, ilyen például a légnyomás, vagy a föld felszínének pontjai. A második csoport a pont adatok, azaz olyan adatok, amelyek csak különböző kitüntetett pontokban léteznek, például földrengések epicentrumai, vagy különböző iskolákból származó adatok. Ezek az adatok lehetnek háromdimenziósak, hiszen pl. a földrengések epicentrumánál lényeges a földrengés felszín alatti mélységei is. Hasonlóan a fertőző beteg ember lakóhelye pontszerű adatnak tekinthető. A harmadik adattípus a területi adat, például olyan adatok, amelyek államokhoz, megyékhez, statisztikai számláló körzetekhez köthetők. Az esetek többségének az ilyen területek valamilyen teljes diszjunkt felosztását adják a vizsgált területnek. Gyakran feltesszük hogy az egyes alapterületi egységek összefüggőek. Minden egyes területi egységhez valamilyen változó értékét rendeljük.

Az adott területi változóink bármilyen mérési szint eredményeként előállhatnak. Gyakran már valamilyen normalizáló tényező segítségével rendezzük értelmes rendbe őket. Gyakran a pont adatokat is úgy kezeljük, mintha azok valamilyen területi egységből származó adatok lennének és az adott területet különböző területi egységekre bontjuk (pl. Thiesseen poligonokra).

Területi adatok általános elemzése

A területi adatok elemzése során valamilyen összefüggéseket keresünk az egyes jelenségek között, vagy az adott jelenség valamilyen sajátosságára szeretnénk rámutatni. Gyakran az a koncepciónk, hogy az adott jelenség eloszlása nem véletlenszerű, hanem valamilyen rendező elvet követ. Az ilyen jelenségeket különböző tesztekkel szoktuk vizsgálni.

Tételek és társadalmi statisztika

A terület statisztikai elemzéséhez szükséges felosztani. Az egyes adatok forrásait, lokalitását meg kell határozni. A különböző tudományágak igen különbözőképpen tekintenek a térre és határozzák meg

azt. (Az építészeti teret például egészen máshogy definiáljuk, mint a statisztikai teret, G.Anker.) A térelemeket nemcsak statikusan, keresztmetszeti szinten elemezhetjük, hanem mintegy dinamikus rendszer elemeit idősorosan sztohasztikus folyamatként. Számunkra ahhoz, hogy területi egységeket tudjunk definiálni, szükséges a határ fogalmának meghatározása. Az irodalomban a határ fogalma négy funkcióval kapcsolódik össze:

- határ, mint elválasztó térelem,
- határ, mint szűrőzóna,
- határ, mint perem,
- határ, mint összekapcsoló elem.

(Például elválasztó elemként szolgálnak az államhatárok, szűrőzóna és kapu szerepkörben vannak a kikötők, peremként értelmezhetőek egy-egy járvány terjedési szintjei, míg összekapcsoló elemként különböző határokon átnyúló speciális gazdasági övezetek tekinthetők.)

Számunkra az elemzéshez a továbbiakban elegendő az, hogy a határok egy adott vizsgált területet diszjunkt egységekre bontanak. Gyakran az így keletkezett térelemeket nem egy poligon hálózattal határolt területi egységnek gondolunk, hanem egy pontot rendelünk az adott poligonhoz és azzal szimbolizáljuk. Például amikor Magyarország településrendszerére gondolunk, gyakran nem 3200 poligonon, hanem 3200 ponttal ábrázoljuk az országot. Kérdés, hogy az adott poligont melyik benne lévő ponttal reprezentáljuk. Erre vonatkozóan különböző koncepciók léteznek az adott poligon középpontjával, esetleges súlypontjával lehetséges a poligont definiálni.

Számít-e a terület?

A szakirodalomban a területi elemzések társadalomstatistikai relevanciáját illetően napjainkban is vita van. Kérdés az, hogy az egyes cselekvő egyénektől függetlenül létezik-e a környezet befolyásoló hatása, azaz létezik-e egy olyan sajátos egyénekre vissza nem vezethető hatás.

Előfordul, hogy az egyének magatartására vonatkozó ismereteink alapján nem tudjuk megfelelően előre jelezni, a vizsgált jelenség alakulását. Egy település bizonyos párt irányába történő választási hajlandósága magasabb lehet, mint amire az ott élő egyének általuk ismert tulajdonságai alapján számíthatunk. Az egyénenként mért magatartás és a közösség mint egész magatartása különbözhet. Így gondolhatjuk, hogy a környezet hatására az adott közösség szintjén olyan új magatartásformák állnak elő, amelyek megértéséhez kevés az egyének magatartásának ismerete, így szükség lehet olyan magyarázó sémák kidolgozására, amelyek a közösség szintjén létező magyarázó elvek felhasználását igénylik. Kérdés, hogy a fenti gondolatmenet elfogadható-e? Ha módszertani individualizmus bázisára helyezkedünk, lehetséges egy alternatív magyarázat is: a környezet módosítja az egyének magatartását, mi azonban nem rendelkezünk erre vonatkozóan elegendő ismerettel, ezért amikor aggregált szinten elemzünk, figyelmen kívül hagyjuk ezt a hatást, és amikor az egyéni cselekvéseket közösségi szinten elemezzük, akkor konstatáljuk az előbb említett hatást. Tehát ha elegendő az adott közösségre, lokalításra vonatkozó ismerettel rendelkezni, akkor már egyéni szinten is konstatálnunk azt a hatást, amit elegendő információ híján csak közösségi szinten észleltünk. Így vizsgálatunkban területiként vagy közösségiként egy olyan jelenséget észlelünk, amely valójában nincs.

Az, hogy melyik magyarázó séma az elfogadható, ez az általunk elfogadott elméleti kerettől függ, megítélésünk szerint azonban amennyiben lehetséges, célszerű az egyszerűbb individualisztikus sémát elfogadni. Érdekes megfigyelnünk azt, hogy ezzel a kérdéssel már több mint száz év óta birkóznak a társadalomtudományok, bizonyos szempontból Durkheim Öngyilkosság c. könyvének is e módszertani probléma áll a homlokterében.

Más nézőpontból megközelítve a fenti problémát, egy adott kutatás esetén a vizsgálatban vannak magyarázó változóink míg a vizsgálat szempontjából külsőséges változókat három csoportba oszthatjuk, vannak ellenőrzött változóink, amelyeket jól el tudunk különíteni és mérni, vannak olyan változóink, amelyeket különböző okok miatt nem tudunk tökéletesen kontrollálni, ezek összekeveredhetnek a magyarázó változókkal, végül vannak olyan változóink, amelyek többnyire a véletlen hibának következményei. Vizsgáljuk a második csoportba sorolt változókat. Számolnunk kell azzal a veszéllyel, hogy a vizsgálat során elhanyagolt tényezők hatását tulajdonítjuk modellünkbe

bevont változóknak. (A társadalomtudományokban mérnöki értelemben vett kísérletek nem folytathatóak, így valójában egy realizációból próbáljuk a paramétereket becsülni.)

Az adott lokalitásban élő egyének tulajdonságai összekeveredhetnek a modellben szereplő változókéval, így a „területi” hatás a hibás változó elkülönítés és mérés következménye. Ez a probléma felhívja a figyelmet arra, hogy egy kutatást mindenképpen meg kell, hogy előzzön a vizsgálat elméleti hipotézisének megalkotása.

Gyakran a környezeti hatásként konstatáljuk a szelekciós mechanizmusok működését. Erre vonatkozóan közismert példa a következő: Oktatásszociológiában közismert tény, hogy az iskolákban azok az alacsony sorból származó tanulók jobb eredményt érnek el, mint sorstársaik, akik jobb színvonalú iskolába járnak. Szokás ezt a környezet azaz a jobb színvonalú iskola húzó hatásával interpretálni, lehetséges azonban egy másik magyarázat is, miszerint eleve motiváltabb családokból származó motivált gyerekek kerülnek ezekbe a „jobb” iskolákba. A fentiek jól rámutatnak arra, hogy ugyanazt a jelenséget két egymást kizáró sémával is magyarázhatjuk.

A területi alapú magyarázatok az egyéni tulajdonságok és a környezet elkülönítésén, illetve elkülöníthetőségén alapul. Jól láthatóan ez a megközelítés rendkívüli módszertani nehézségekbe ütközik. A statisztikai elemzés során a változóra nézve is aggregáljuk adatainkat, ez az aggregáció azon az alapfeltevésen nyugszik, hogy a mérés alapjául szolgáló egyedek tökéletesen homogének, hiszen ellenkező esetben a már elemzett téves hatást fogjuk számszerűsíteni. Így nagyon gyakran nem vagyunk képesek megfelelően mérni az elemzéshez szükséges változókat.

Összefoglalásként megállapíthatjuk, hogy a feltett kérdésre - számít-e a terület - megnyugtató választ nem tudunk adni. A kérdést egyszerű módon és általánosan nem lehet lezárni. Konkrét kérdések vizsgálatakor először mindig pontosan rögzítenünk kell vizsgálatunk hipotézisét, és azt az elméleti keretet, amelyben vizsgálatunk során eljárunk. Meggyőződésünk szerint – már csak Occam borotvája okán is – érdemes a módszertani individualizmust, mint vizsgálati keretet elfogadni, bár gyakran tudomásul kell vennünk, hogy ebben az esetben nem jutunk kielégítő eredményre, hiszen egy individuális magyarázatokon felépülő modellhez nem lesz elégséges adatunk. Így a modell specifikálása során detektálni fogunk környezeti hatásokat, melyek újabb vizsgálat tárgyát kell, hogy képezzék az egyéni magyarázó sémára való visszavezetés érdekében.

Társadalomtudományi vizsgálatok során gyakran csak összesített aggregált adatok állnak rendelkezésünkre. Amennyiben a módszertani individualizmus feltevéseit fogadjuk el, azaz úgy gondoljuk, hogy magyarázataink végén a cselekvő egyéneknek kell állnia, és a cselekvő egyén magatartását kell megmagyaráznunk, akkor gyakorta jelentős nehézségekkel kell szembesülnünk. Gondoljuk csak meg, a választási eredmények magyarázatakor csak szavazóköri eredmények állnak rendelkezésünkre (egy-egy szavazókörben 600-1200 választásra jogosult szavazhat), esetlegesen az adott szavazókörre, településre vonatkozóan más társadalomstatisztikai adatok is rendelkezésünkre állnak, így kor, nem, foglalkozás, stb.-re vonatkozó megoszlások. Szavazni azonban egyének szavaznak, így az egyes egyének magatartását kell értelmeznünk. Az egyes egyének magatartására vonatkozó elemzések azonban a fenti adatokból nem, vagy csak részlegesen lehetségesek.

Vizsgáljunk egy egyszerű példát. Hasonló társadalomstatisztikai megoszlásokkal rendelkező településeken lakó szavazópolgárok egészen más választási magatartást tanúsíthatnak. Így az adott terület valamilyen értelemben magyarázatul szolgál elemzésünkben. Nyilvánvaló, hogy vizsgálatunknak többek között arra kell irányulnia, hogy ezt a területi hatást képesek legyünk konkrét hatásokra felbontani, és magyarázni az esetleges társadalomtörténetből adódó változókat. Így vélekedésünk szerint a terület mint magyarázó változó jelentős segítséget adhat a társadalomtudományi magyarázatok terén, azonban tisztában kell lennünk azzal, hogy a terület, mint magyarázó változó nem elégséges, hanem további hipotézisek, magyarázatok, vizsgálatok szükségesek.

Így összefoglalásként álláspontunk köztesként tekinthető az irodalomban elterjedt két álláspont között. Az egyik álláspont (King 1996) szerint a területi kontextus, mint magyarázó változó elfogadhatatlan a társadalomtudományi magyarázatokban, még munkahipotézisként sem. A másik álláspont (Agnew 1996) szerint a területi kontextus gyakran fontos magyarázó tényező lehet a társadalomtudományi

magyarázatokban. Hazánkban a társadalomtudományi területi statisztikával foglalkozó szakmai publikációk száma igen csekély, így az olvasó figyelmébe csak Moksony Ferenc Kontextuális elemzés c. publikációjának idevonatkozó bekezdéseit ajánlhatjuk az olvasó figyelmébe. Könyvünkben a területi statisztika és gyakorlati alkalmazásainak legfontosabb aspektusait tekintjük át, részben politikai választási, részben humán epidemiológiai példákon nyomán követve az egyes módszereket.

Ökológiai hatások és területi kapcsolatok

A kérdés vizsgálata a második világháború után kapott új lendületet, Leo Goodmann munkássága nyomán, aki a weimari Németország szavazói magatartását vizsgálta. Az 1930-as németországi választásokon a korábban viszonylag jelentéktelen náci párt a voksok 18,3%-át kapta, mely arány 37,3%-ra növekedett az utolsó választáson. A kortársak közül is, de a második világháború után is sokan kérdezték, hogy kik szavaztak a náci pártra, a lecsúszó középosztály, esetleg a protestánsok, esetleg a katolikusok, netán az értékesítési nehézségekkel küszködő gazdálkodók, stb. A kérdésre nem egyszerű a válasz, hiszen a választási eredmények választó körzetenként állnak rendelkezésre és hasonlóan a népszámlálási adatok is ilyen bontásban állnak rendelkezésre, miközben a szavazás egyéni cselekvés. Így a problémánk a következő: az összegzett adatokból szeretnénk az egyes egyének magatartására következtetni. Lehetséges-e ez egyáltalán?

A kutatásoknak újabb lökést adott az Egyesült Államok választási törvényének 1965. évi módosítása, mely megtiltotta az etnikai hovatartozás szerinti diszkriminációt. Amennyiben ez mégis előfordulna, a bíróság elrendelheti a szavazási körzetek újraalakítását. A törvény szerint diszkrimináció akkor keletkezik, amennyiben az adott etnikai kisebbség az adott területen a többi szavazótól egyértelműen különböző módon szavaz, és a többségi szavazók megakadályozzák a kisebbségi szavazókat állampolgári jogaik érvényesítésében. A fenti tényállás bizonyítása igen nehéz, hiszen választási adatok az Egyesült Államokban is (mint bárhol a világon, ahol titkos szavazási jog van) csak szavazókör szinten állnak rendelkezésre. Tehát újra az ökológiai probléma megoldásához jutunk.

A kérdésre több megoldási javaslat született, Goodman (1953) regressziót, Duncan és Davis (1953) a határok módszerét javasolta. A két legújabb modell King (1997) és Wakefield (2003) bayesi alapú likelihood-bebecslést alkalmaz.

A kutatás részeredményeit közlő írásunk: Mészáros, Molnár D., Mérő Az aggregált adatok dekomponálása, Szociológiai Szemle) a szakirodalomban elterjedt és használatos módszereket tekinti át, illetve egy adatstruktúrán bemutatja azok alkalmazását és a módszerek korlátait.

A közvéleménykutatási adatok esetében jól meghatározott mintavételi terv kialakítása szükséges. E mintavételi terv segítségével történő lekérdezés során individuális adatok nyerhetők. Így lényegében két mátrix áll rendelkezésre. Az egyiknek csak a peremei ismertek, míg a másik teljeskörű. A teljeskörű mátrix segítségével megkísérelhető bayesi becslés adása a részlegesen ismert adatstruktúrára.

Az utóbbi fél évszázadban (Robinson 1950, Goodman 1953) számos publikáció és vizsgálat született az ún ökológiai probléma megoldására, azaz arra, hogy az aggregált adatokból milyen módon tudunk a cselekvő egyénekre következtetéseket levonni.

A probléma kettős. Két típusú területi hatás lehetséges. Az egyik az ún területi autokorreláció, a másik a területi keveredés. A két eset különböző módszertanokat igényel. Vizsgáljuk először a területi keveredést.

Az ökológiai probléma első megoldási kísérletét (Goodman 1953) az a feltevés jellemezte, hogy a különböző területi egységek (melyek az adatforrásaink) homogénnek tekinthető, azaz az egyes területi egységekben az egyének megoszlása állandó. E feltevés nyilvánvalóan túl erős, a gyakorlatban tarthatatlan. Így Goodman megoldási módszere a gyakorlati alkalmazás próbáit nem állja ki. Az azóta született megoldási kísérletek mindegyike különböző többé-kevésbé erősen kritizálható feltevésekkel él a területi egységek homogenitásáról. (Az irodalom jó áttekintését adja Wakefield 2002.) A gyakorlati életben a területi homogenitás nagyon kevésbé tartható hipotézis. A településtudományok

egyik ma már közhellyé vált megállapítása a szegregáció jelensége, azaz, hogy különböző területi települési egységek rendkívül inhomogén népességet tartalmazhatnak.

A területi elemzések valamilyen területi egységre vonatkozó (települési, szavazóköri, számláló körzeti) adatokra hagyatkoznak. Gyakorta azonban ezek az egységek igen kevésbé alkalmasak az érdemi elemzésre, hiszen ezek vagy történeti, vagy közigazgatási hagyományok alapján jöttek létre, melyek gyakran kevésbé relevánsak az adott elemzés számára. Gyakran például az adott területen élő munkanélküliek számát és valamilyen párt támogatottságát kíséreljük összemérni, azonban az így kapott korrelációs együttható igen megtévesztő lehet. (A későbbiekben mutatunk ilyen példákat.)

A területi adatok legfontosabb jellemzője az autokorreláció, ami jelen vizsgálatunk számára pusztán annyit jelent, hogy az adott jelenség nem követ egyenletes megoszlást az adott területi egységek között, azaz a jelenség tömbösödik. A területi jelenségek vizsgálata két csoportra osztható. Az egyiket hagyományosan területi statisztikának szokás hívni, itt területi pont folyamatokat vizsgálunk, míg a másik módszertant területi ökonometriának szokás nevezni, itt többnyire területi adatok regressziós vagy többváltozós elemzése zajlik. Az első esetben többnyire valamilyen általánosított térképi felületről, pontokból álló mintát veszünk, és ezek képezik későbbi vizsgálataink alapját, még a második esetben a területi egységekről származó adataink közötti valamilyen összefüggés rendszert kísérelünk kimutatni.

A kontextuális hatás és vizsgálata

A területi elemzések alapvető különbségek, közgazdasági, politológiai vagy szociológiai elemzésektől, az adatok és elemzések kontextuális jellege. A politikai elemzések általában nem szokták vizsgálni és tárgyalni annak a területnek a hatását az adott politikai folyamatra, amely területen az végbement. (A járványtani vizsgálatoknál jellemzően ez a vizsgálatok egyik legfontosabb eleme, hiszen a környezeti vizsgálatok nélkül elképzelhetetlen egy járványtani elemzés.

A földrajztudományok képviselői többnyire amellet érvelnek, hogy a területi hatás számít és a vizsgálatok során nem hanyagolható el, még a társadalomtudományok képviselői többnyire úgy érvelnek, hogy ilyen hatás vagy nincs, vagy megtévesztő, hiszen valójában nem a területi hatásról van szó, hanem különböző nem számszerűsített magyarázó hatások összeadódásáról, melyet a földrajztudósok kontextuális hatásnak neveznek.

Mi a kutatásunk során köztes álláspontot foglaltunk el, azaz számunkra fontos dolog volt a területi hatás számszerűsítése, még ha tudjuk is, hogy ez pusztán vizsgálataink egy köztes állomása kell, hogy legyen.

Az egyéni magatartás és az aggregált adatok

Az előző okfejtésből világos, hogy vizsgálataink centrális kérdése volt az, hogy hogyan tudjuk összekapcsolni adatainkat, melyek területi szinten állnak rendelkezésre, a cselekvő egyének magatartásával.

Az aggregált adatokból való következtetések mint az közismert, gyakran téves eredményekre vezetnek. (Úgynevezett ökológiai tévkövetkeztetés, vagy Simpson paradoxon, a kutatás során tanulmány is íródott e tárgyban) Így vizsgálatainkban nem hagyatkozhatunk az aggregált adatok egyszerű manipulálására. Az irodalomban hagyományos vizsgálati kérdés különböző módszertanok számára mintegy kályhával szolgálnak az 1930-as évek német választási adatai, azaz ki szavazott a náciakra.

Nem-stacionárius adatok és a területi autokorreláció

Nagyon gyakran a területi elemzések során az egyes népességi változók közötti kapcsolatot keressük, azaz a fontos területi vagy regionális körülményeket szeretnénk a vizsgált körülmény magyarázatához felhasználni. Például az egyes területek népességének szavazási eredményeinek értelmezéséhez más

forrásokból származó társadalomstatistikai, demográfiai adatokat. Azonban ezek az adatok nagy szóródást mutathatnak a különböző területi egységek között (szakszóval mondván klasztereződnek). Így módon a globális és lokális mutatók azonos elemzési keretben történő felhasználása nem probléma mentes, mivel előző példánknál maradván az adataink klasztereződtek, más szakszóval mondván nem stacionáriusak (azaz az adott jellemző várható értéke a különböző területi egységekre vonatkozóan nem állandó). Megjegyezzük hogy a nem-stacionárius adatok elemzéséhez a standard módszertan nem jól használható.

Szomszédossági hatások

A társadalomtudományi irodalomban mindmáig nem nyugodott meg a vita a szomszédossági hatások jelentőségét vagy létezését illetően. (Amennyiben a módszertani individualizmus gondolati rendszerében kívánjuk tudományunkat felépíteni, nyilvánvalóan a szomszédossági hatás, mint magyarázó tényező önmagában nem elegendő). Mi, mint a tanulmányban a korábbiakban is, köztes álláspontot foglalunk el. Ugyan a módszertani individualizmus feltevésrendszerét fogadjuk el, de ettől még megítélésünk szerint a területi hatások és szomszédossági hatások számszerűsítését lehetségesnek és fontosnak tarthatjuk, azzal a megszorítással, hogy e hatások keletkezése és hatásmechanizmusának felderítése további elemzést igényel. Az utóbbi években két jelentősebb összefoglaló is készült e tárgykörben, Robert D. Dietz (2002) illetve George C. Galster (2008). Manski (1993-2000) a szomszédossági hatásokat 3 nagy csoportba sorolta: az endogén, a kapcsolódó és az exogén hatások csoportjába. Az endogén hatások egy adott szomszédosságon belül jelentkeznek, amikor a szomszédosság egy tagja közvetlen hatást gyakorol a többiek magatartására. Az endogén hatások közé szoktuk sorolni a „Bandwagon” és a személyes (Peer) hatásokat. Ezekben az esetekben az oksági lánc, amely ezt a hatást kiváltja jól nyomon követhető és így kimutatható. Az ilyen hatások gyakorta többszöröződnek, azaz visszacsatolási (autoregresszív) mechanizmus van jelen. Vizsgáljuk meg a következő egyszerű esetet. Egy adott környezetben tegyük föl, hogy a fiatalok drogfogyasztása függ attól, hogy a többi fiatal milyen gyakori drogfogyasztó. Ebben az esetben, ha akárcsak egy fiatal sikerül drogmentesség tennünk, ez hatást gyakorol az összes többi fiatalra és jelentősen csökkenti a drogfogyasztás összszintjét az adott környezetben. Kapcsolódó (korrelált) hatásról akkor beszélünk, amikor az adott környezetben élő egyének valamilyen oknál fogva hasonló magatartást igyekeznek felvenni. Ilyen jelenség például egy adott lakókörnyezet homogenizáltsága, azaz a szegregáció jelensége, amikor hasonló státusú emberek laknak egy adott lakókörnyezetben. Az endogén jelenségekkel ellentétben a kapcsolódó jelenségek számszerűsítése nem egyszerű. Megjegyzés: az oksági és nem oksági környezeti hatások elkülönítése igen fontos. Hiszen a két különböző típusú hatásnak különböző elméleti következményei vannak. Tegyük fel például, hogy a drogfogyasztás fordítottan arányos a jövedelemmel. Tegyük fel továbbá azt is, hogy a jövedelem csoportosító hatást gyakorol a lakókörnyezetekbe költözésre nézve. Így a szegénység koncentrálódása egyes lakókörnyezetekben, melyet a lakáspiac működésmechanizmusa okoz, a drogfogyasztás koncentrálódását is eredményezi. Ebben a példában a drogfogyasztás egy nem oksági korrelált hatás. Társadalmi multiplikátorok sem keletkeznek ebben a megközelítésben, hiszen a magatartás nem a környezeti hatás eredménye.

Exogén hatások (az angol nyelvű szakirodalom gyakran „kompozicionális hatásnak” nevezi ezeket). Ilyen hatások akkor jelentkeznek, amikor az egyének a szomszédok vagy a szomszédosság valamilyen exogén jellemzőitől függenek. Így például egy-egy adott lakókörnyezet faji, vagy vallási összetétele. Ilyen esetekben a hatások számszerűsítése viszonylag egyszerű, hiszen csak az exogén változók közötti kapcsolatot kell számszerűsíteni.

Érdemes megjegyeznünk azt is, hogy a szomszédosságok között is létezhetnek hatások. Ilyenek például az úgynevezett externáliák. Ebben a megközelítésben a szomszédosságoknak hatása van egymásra (ekkor az egyes környezeteket nem tekintjük független változóknak, hanem a kölcsönös összefüggéseket kíséreljük meg számszerűsíteni).

Az előzőekben tárgyalt meghatározások keresztmetszetiek voltak, azaz eltekintettünk az időtényezőtől e jelenségek időbeli vizsgálatától, lefutásától. A szakirodalomban lévő elemzések (valószínűleg az adatok hiánya miatt) nem nagyon tartalmaznak ilyen típusú vizsgálatokat.

A környezeti hatások elméleti megközelítései

Terjedési megközelítések.

Ezekben a vizsgálatokban az egyes társadalmi jelenségek terjedését vizsgáljuk döntően a személyközi kapcsolatok útján. A népegészségügyben szokásos epidemiológiai vizsgálatok tipikusan ilyen terjedési folyamatokat vizsgálnak, amikor egy adott környezetben egy betegség megjelenését és feldúsulását vizsgáljuk.

Másik szokásos megközelítés a szocializációs modell. Ebben az esetben valami a társadalmilag hasznos magatartás elterjedését az egyének közötti interakcióknak és a közösségi hálózatoknak tulajdonítjuk.

A harmadik modell típus a verseny. Ebben az esetben úgy tekintjük, hogy az adott jelenség szempontjából vannak győztesek és a győzteseknek hatásuk van az adott közösség többi tagjára. Az előző megközelítésnek ellentettje a relatív deprivációs elmélet, amely azt mondja, hogy a közösség egyes tagjainak pozitív emelkedése negatív hatást gyakorolhat a szomszédokra. Ebben a megközelítésben a „győztesek” nem magukkal vonzzák a többieket, hanem erősítik a többiek leszakadását.

A környezeti hatások számszerűsítése

A környezeti hatások számszerűsítésével számos vizsgálat, modell született. Meg kell jegyeznünk azonban, hogy e modellek felírása, számszerűsítése döntő mértékben függ attól, hogy milyen elméleti keretben gondolkodunk. Az alábbiakban egy áttekintő tipológiát közlünk.

1. Endogén hatások: a környezetben élők egyedeinek más egyedekre gyakorolt közvetlen hatása. Ilyen hatások felléphetnek:

- szocializáció: amikor az egyes egyedek magatartása a lakókörnyezetben lakók magatartásának közvetlen hatására alakul ki
 - normák: a szocializáció speciális esete, amikor a magatartás bizonyos környezeti kényszerek hatására alakul
 - szelektív szocializáció: amikor ugyanaz a hatásmechanizmus különbözőképpen hat az egyes egyedekre, például iskolai jutalmazás, büntetés stb.
 - hálózatok: a szocializációs mechanizmus konkrét lefutását is be tudjuk mutatni az egyes egyének közötti kapcsolatokon keresztül
- verseny: amikor a helyi erőforrások korlátozottak, és az adott környezetben élők versenyeznek ezen erőforrásokért
- relatív depriváció: a korlátos erőforrásokért való harc nem csak győzteseket, de veszteseket is létrehoz
- stigmatizáció: a verseny nem csak anyagi, de szimbolikus erőforrásokért is történhet
- externális hatások: az adott lakókörnyezetben történő változások tovaryűrűző hatást is okozhatnak

2. Kapcsolódó hatások: az ilyen hatások esetén valamilyen jelentős külső hatáshoz kapcsolódik a környezet illetve a környezeten belüli változások

3. Külső hatások: ebben az esetben a hatásokat valamilyen olyan valamilyen exogén változó hatásának tulajdonítjuk pl. vallás, faj. Például az újbevándorlók szeretnek olyan lakókörnyezetbe költözni, ahol többnyire hasonló nemzetiségűek élnek

A fenti egyenlet paramétereinek meghatározása igen kiterjedt adatgyűjtést igényel, amely gyakorta nem lehetséges. A szomszédsági hatások számszerűsítése azonban igen hasznos is lehet, hiszen a vizsgálataink során a pontos egyes egyénekre ható hatásmechanizmust nem is tudjuk számszerűsíteni, de a társadalmi jelenségek összekapcsolódását, területi összefüggéseit igen, amely tudományosan ugyan lehet, hogy nem kielégítő, de a gyakorlati alkalmazások szempontjából sok segítséget ad. Így például bizonyos társadalmi jelenségek koncentrációját és makroszintű hatásmechanizmusait lehetséges számszerűsíteni, és az esetleges beavatkozások következményeit is lehetséges előrejelezni. Hasonlóan, mint azt a területi mintavételezésnél megfigyelhetjük, jelentős erőforrások takaríthatók meg, amennyiben a vizsgálat tárgyát képező jelenség belső területi összefüggéseit figyelembe vesszük a mintavételnél is.

Mintavételezés és területi statisztika

Kutatásunk másik fókusza a területi szempontú mintavételezés volt. A jelenleg a társadalomkutatásban elterjedt mintavételező eljárások mindegyike feltételezi az adott mintastruktúrában a területi homogenitást. (Az egyszerű véletlen mintánál ez szükségszerű feltevés, rétegzett mintavételezés esetén a rétegeken belüli valamilyen értelemben vett homogenitást kell feltételeznünk.) E feltevés a gyakorlatban nem teljes. A területi adatok autokorreláltak, azaz területi blokkosodást mutatnak és gyakorta az irányhomogenitás sem teljesül (anisotropikusak). A műszaki tudományokban ugyan kis területeken, de a fenti feltevéseket nem alkalmazó mintavételi eljárások alakultak ki. Megítélésem szerint érdemes lenne ezen eljárások társadalomstatisztikai adaptálása.

A mintavételezési eljárás, melyet gyakorta valamilyen volumen becslés követ, felhasználja a területi egységekről már rendelkezésre álló információt. A mintavételi keret kialakításánál figyelembe veszi az „adott réteg” vastagságát. Így a mintavétel elemszáma („fűrészek” száma) jelen esetünkben a mintába kerülő területi egységek száma csökkenthető és a volumen becslés pedig megbízhatóbbá válhat. (A jelenleg használatos mintavételi technikák esetében ugyan a sajtó publikációban 3%-os mintavételi hiba szerepel 1000 fős minta esetében, szeretném hangsúlyozni, hogy ez a megállapítás gyakorta téves, hiszen a mintavételi hiba számításának feltevései nem vagy csak részben teljesülnek, mely feltételek ellenőrzése gyakorta elmarad.)

A matematikai statisztika széleskörű, gyakorlati alkalmazásai mint egy 150 éve kezdődtek (tekintsünk most el Gauss idevonatkozó munkásságától). A statisztikai megközelítés lehetővé tette a teljes sokaság vizsgálati helyen, annak egy részének vizsgálatát. Ez a megközelítés felveti annak kérdését, hogy ezt a részt hogyan is kell meghatározni. Ez a mintavételezés kérdése. A mintavételezéssel kapcsolatosan számos kézikönyv látott napvilágot. Ezek többé-kevésbé közismertek, így az érdeklődők figyelmébe ajánlom Kish, Cochran könyvét. Ezek a könyvek a standard megközelítéseket tartalmazzák. A standard megközelítések jó összefoglalója olvasható Rao (1999) cikkében. Ez a cikk napjainkig összefoglalja a statisztikai mintavétel legfontosabb irányait. Hasonlóan nemrégiben elhunyt honfitársunk Leslie Kish utolsó írásai is lényegében az utóbbi 150 év legfontosabb folyamatait foglalják össze. Ezek az írások azonban megítélésem szerint két fontos megközelítést nem tárgyalnak. Az egyik a mintavétel területi aspektusai, a másik meg a bayes-i statisztikai megközelítés. Mi a következőkben e két megközelítéssel kívánunk foglalkozni és megkísérjük e két megközelítést összekapcsolását.

A mintavételezés az esetek többségében (társadalomtudományok esetében döntő részében) valamilyen területi egységhez kötődik. Mit is jelent ez? Amikor valamilyen társadalmi jelenséget kívánunk vizsgálni, politikai preferenciákat, sörfogyasztási szokásokat... stb., akkor az esetek többségében viszonylag kis esetszámú, mondjuk 1000 fős rétegzett mintával fogunk dolgozni. Azaz elhatározzuk (döntően költségvetési okok miatt), hogy mondjuk 1000 fő vizsgálatára nyílik lehetőségünk, és ezt az 1000 főt kívánjuk optimálisan allokálni az adott területen. Többnyire ebben az esetben többszempontú rétegzést hajtunk végre. Mondjuk Budapesten lekérdezzük 200 főt, az 5 „nagyvárosban” 100 főt és a különböző településtípusok rétege méretének megfelelően folytatjuk ezt az eljárást. Ebben az esetben avval a feltevéssel élünk, hogy az adott jelenség az adott rétegen belül homogén, azaz teljesen mindegy, hogy az a település Nyugat-Dunántúlon, a Tiszántúlon van, lakói ugyanolyan magatartást

követnek az adott jelenség szempontjából. Ez gyakorta nem igaz, így e homogenitás feltevése jelentős hibát eredményezhet. Éppen ezért a mintavétel területi aspektusaival mindenképpen foglalkoznunk kell, hiszen az az ismeret, hogy az a jelenség az adott értelemben nem homogén, többnyire rendelkezésünkre áll.

Bayes-i statisztika

Napjaink statisztikájában egy csöndes bayes-i forradalom zajlik. Míg a 20.század egyértelműen a klasszikus Kolmogorov-féle gyakoriság-mértékelméleti megközelítés uralta, napjainkban változik a szemléletmód. Sokan újra vizsgálják a bayes-i megközelítést, rámutatva annak számos gyakorlati eredményére. (Könyvünk végén egy külön fejezet foglalkozik a bayes-i statisztika bemutatásával.)

A bayes-i megközelítésben, ha valamelyes ismeretünk rendelkezésre áll a jelenség korábbi lefutásáról, azt jól tudjuk hasznosítani a mintavételi terv kialakításakor.

Bevezetés Bayes-i statisztikába

A statisztika-elmélet kialakulásának idején az 1700-as 1800-as évek fordulóján a valószínűség fogalmának szigorú értelemben vett tudományos megalapozása hiányzott. Napjainkból visszatekintve azt mondhatnánk, hogy a szubjektív valószínűségű és a gyakorisági iskola nézetei keveredtek. Az 1800-as évek második felében Laplace a szubjektív valószínűség fogalmát definiálta és erre a fogalomra építkezve dolgozott ki egy valószínűség-elméletet. E valószínűség-elmélet napjainkig élő hagyománnyá vált. Az 1900-as évek elején Kolmogorov majd Neyman munkássága nyomán kidolgozásra került a „klasszikus iskola”, melyet szoktak gyakorisági megközelítésnek is nevezni. A XX. század lényegében a klasszikus iskola diadalútja volt. Napjainkban azonban egy „bayes-i forradalmat” élünk át. Ha megvizsgáljuk a publikációkat, akkor azt tapasztaljuk, hogy egyre több cikk, támpont születik bayes-i szellemben. Ebből a nemzetközi tudományos tendenciából a tanulmányok tanulsága szerint hazánk kimarad, bayes-i szellemű cikkek, könyvek nem nagyon születnek, nem kerülnek publikálásra. Magyar nyelven mindössze egy (németből fordított) bevezető szakkönyv olvasható. Megítélésünk szerint hasznos lenne ha a hazai szakirodalom is kiegyensúlyozottabbá válna. Nem akarjuk vitatni a gyakorisági iskola nagyon jelentős eredményeit, de látnunk kell ennek az elméletnek a korlátait is. Ahogy hasonlóan a bayes-i statisztika sem csodaszer, vannak jelentős előnyei és fogyatékososságai. Érdemes áttekinteni a két valószínűségi iskola koncepcionális különbségeit röviden, nagyon elemi módon.

A bayes-i statisztika fogalmai

- 1.) A priori valószínűség: annak a valószínűsége, hogy a vizsgált modell igaz, még az adatok észlelése előtt.
- 2.) Posteriori valószínűség: annak a valószínűsége, hogy a vizsgált modell igaz, a figyelembe vett adatok fényében.
- 3.) Likelihood: egy adott modell választása esetén az adatok feltételes valószínűsége.

A bayes-i statisztika fontos tulajdonsága, hogy csak a ténylegesen megfigyelt adatokat vesszük figyelembe annak eldöntésére, hogy a vizsgált modell helyes-e.

A valószínűség fogalma

A korábbiakban olyan kifejezéseket használtunk, mint „annak a valószínűsége, hogy a jó érmét dobáltuk fel”, a p valószínűség eloszlása stb., ezek a kifejezések a gyakorisági iskola fogalmi rendszerében értelmetlenek például a p konstans így nem lehet eloszlása stb. ...

Ha a klasszikus iskola keretei között gondolkodunk, akkor, ha kiválasztunk egy érmét, akkor p a továbbiakban már nem véletlenszerű. Az érme kiválasztása és ekként a hipotézis rögzítése értelmetlenné teszi az olyan kijelentéseket, mint „annak az esélye, hogy a hamis érmét választottuk...”. Az ilyen típusú kijelentések személyes bizonytalanságot fogalmaznak meg és így szubjektív mértéket jelentenek.

A gyakorisági megközelítés szerint a valószínűség hasonló körülmények között lefolytatott kísérlet sorozatból adódó relatív gyakoriság. Ha egy esemény már megtörtént, akkor már nem véletlenszerű, így nem lehet a valószínűségről beszélni.

Ezért van az, hogy a konfidencia intervallum interpretációja igen nehézkes. Ha azt állítjuk, hogy a 95% az esélye annak, hogy a paraméter a konfidencia intervallumba esik, akkor rögzített (nem valószínűségi jellegű) dolgokhoz rendelünk valószínűséget. Helyesen ez így hangzana: ez az intervallum egy olyan eljárás eredménye, mely az esetek 95% -ban olyan intervallumot eredményez, mely p -t tartalmazza.

A valószínűség szubjektív (bayes-i) értelmezése szerint a valószínűség egy adott személy ítélete az adott eseménnyel kapcsolatban. Így ez a megközelítés – meglátásunk szerint – szélesebb körben alkalmazható, egyszeri események, múltbeli események stb. Ebből következően az egyes egyének különböző valószínűségeket rendelhetnek az eseményekhez, hiszen helyzetük, informáltságuk különböző lehet.

Diagnosztikai teszt eredményének értelmezése a Bayes-tétellel

Valaki tavaly elment és tetováltatta magát, 2 hónappal később elment vért adni, de nem engedték. Azt mondták neki, hogy 1 év múlva jelentkezzen legközelebb véradásra, hogy biztosak lehessenek abban, hogy nem kapott hepatitis B fertőzést a tetoválástól. A fiatalember elkezdett aggódni és vett egy házilag alkalmazható tesztet a hepatitis B vírussal (HBV) történt fertőzés kimutatására. A tesztről azt olvasta a tájékoztatóban, hogy 99%-os a szenzitivitása és 99,5% a specificitása. A szenzitivitási érték azt jelenti, hogy 100 fertőzött emberből 99 esetében kimutatja a teszt a fertőzöttséget. A specificitásra vonatkozó érték pedig azt jelenti, hogy a megvizsgált nem fertőzött emberek 99,5%-ában negatív eredményt ad a teszt, azonban 0,5%-uknál téves pozitív eredményt. A tájékoztatóban az eredmény értelmezésére vonatkozóan is kapott információkat. Ezek szerint ha negatív a teszt eredménye, akkor az azt jelenti, hogy nem fertőzött a vírussal. Ha viszont pozitív a gyors teszt eredménye, nem jelenti teljes bizonyossággal azt, hogy fertőzött a vírussal. Ha a teszt eredménye pozitív tanácsos orvoshoz fordulni.

A hepatitis B egy ritka betegségnek számít azok között, akik nem használnak intravénás drogokat, ebben a populációban nagyjából 2 ember fertőzött 100,000-ből. Több tanulmány is azt mutatta, hogy a megfelelő higiéniajú tetoválás során nem növekszik az esély a fertőződésre. Gondolhatjuk azt, hogy azok között, akik tetováltatják magukat, eltérhet a fertőzöttség mértéke, mondjuk magasabb, tegyük fel, hogy 3/100,000.

A Bayes-tétel felhasználásával kiszámíthatjuk, hogy mekkora annak az a posteriori valószínűsége, hogy fertőzött, ha a gyors tesztje pozitív eredményt mutatott. Állítsunk fel két modellt, $M1$ = a fiú pozitív és $M2$ = negatív HBV-re vonatkozóan. Az előző információk alapján az apriori valószínűségek $P(M1) = 3/100,000 = 0,00003$ és ebből következően $P(M2) = 0,99997$.

A likelihoodok meghatározásához idézzük fel, hogy mit is értettünk alatta: annak a valószínűségét, hogy a megfigyelt adatok a modellből származnak. Összesen kétféle érték lehetséges a teszt eredményeként: pozitív vagy negatív teszt. Először vegyük szemügyre az első eredményt (D) leíró mondatot: „A fiú teszt eredménye pozitív.” Ebben az esetben mindkét modellre ki kell számítanunk a likelihoodot arra vonatkozóan, ha a teszt pozitív és ezután számítható Tamás fertőzöttségének a posteriori valószínűsége.

Ha az $M1$ modell igaz (fiú fertőzött, de nem tud róla), akkor 99% annak a valószínűsége, hogy pozitív a teszt eredmények. Ezt onnan tudhatjuk, hogy a teszt szenzitivitása 99%-os, ami azt jelenti, hogy a fertőzöttek 99%-ában pozitív eredményt fog adni. Következésképpen az $M1$ modell likelihoodja $P(D | M1) = 0,99$. Másrészről ha az $M2$ igaz (fiú negatív HBV-re), akkor 99,5% a valószínűsége, hogy a teszt negatív eredményt ad, vagyis csupán 0,5% a valószínűsége annak, hogy pozitív eredményt

kapunk. A teszt eredményétől függően számítjuk a likelihoodot, ha a teszt eredménye pozitív vagy ha negatív.

Ha a teszt eredménye negatív, annak számítását az alsó tábla mutatja. Az apriori valószínűségek azonosak, mivel azok a fiú véleményét tükrözik (mindkét esetben) a teszt eredményének megismerése előtt. Mivel a D adat (a teszt eredménye negatív) más, ezért a likelihood is más lesz. Az $M1$ modell esetén a likelihood – ha negatív a teszt eredménye – annak a valószínűsége, hogy egy fertőzött ember negatív eredményt ad (téves negatív), ami $1 - \text{szenzitivitás} = 0,01$, vagyis 1%. Az $M2$ modell esetén a likelihood annak a valószínűsége, hogy egy negatív ember tesztje negatív eredményt ad, vagyis a specificitás, ami itt 99,5%. Annak a valószínűsége, hogy negatív teszt eredmény mellett a fiú fertőzött, nagyon jelentősen lecsökkent, $3/10.000.000$ ennek a valószínűsége.

Fontos megjegyezni, hogy az apriori valószínűségek nem függnak a megfigyeléstől, mivel azok az új információk előtti véleményünket fejezik ki. A megfigyelt adatok hatása a véleményalkotásunkra teljes egészében a likelihoodon keresztül valósul meg.

A klasszikus iskola megközelítése szerint a vizsgált „beteg”-ről valószínűségi kijelentést tenni értelmetlen, hiszen vagy hordozza a betegséget, vagy nem ez determinisztikus. A bayes-i megközelítés szerint ez értelmes kijelentés

Konfidencia intervallum

A gyakorisági iskola konfidencia intervallumot használ a paraméter becslésére. A konfidencia intervallummal kapcsolatos valószínűségi kijelentés a mintavételi eljárásra vonatkozik. A többször ismételt véletlen mintavételi eljárás olyan intervallumot eredményez, mely az esetek 95%-ban (az előre megválasztottak szerint) tartalmazza a paramétert. Ha a mintavétel megtörtént, a konfidencia intervallumot kiszámoltuk, már semmi sem véletlenszerű, nincs helye valószínűségi kijelentéseknek.

A bayes-i megközelítésben is használhatunk intervallumot a paraméter becslésére, ekkor ezt hihető (credible) halmaznak, intervallumnak nevezzük. Ez az az intervallum, melyben a paraméter p valószínűséggel van.

Az a priori eloszlás választása

A bayes-i megközelítést gyakran éri kritika, hogy az a priori eloszlás önkényes és így az eljárásban sok a szubjektív elem. Ez igaz, de ez az állítás igaz a gyakorisági iskolára is, hiszen a hipotézis felállítás a szignifikancia szint is önkényes lehet. A minta növelése azonban csökkentheti az a priori eloszlás választásának hatását.

Összefoglalás:

Az alábbi táblázatban a két megközelítés különbözőségét mutatjuk be.

	Gyakorisági	Bayes-i
P	a hasonló körülmények közötti kísérletekből adódó relatív gyakoriság	a megfigyelő valószínűségi ítélete az adott eseményről
statisztikai kapcsolat	a feltett modell esetén a tapasztalt adatok valószínűségi kiértékeléséből adódik	a megfigyelt adatokból a vizsgált modell valószínűségi kiértékeléséből adódik
konfidencia intervallum	annak a valószínűsége, hogy egy előre meghatározott érték szerint a követett eljárás az adott valószínűséggel fog olyan intervallumot eredményezni mely az adott paramétert tartalmazza	
Hihető (credible) intervallum		Az adott meghatározott valószínűséggel a paraméter az intervallumban van
Szignifikancia teszt P értéke	annak a valószínűsége, hogy amennyiben a Null-hipotézis igaz, akkor a vizsgált paraméter megfelelő	Az adott adatok mellett értékeljük a modellt. (a nem megfigyelt adatok érdektelenek.

A kutatáshoz kapcsolódó publikációk

Könyv:

Mészáros József: Elméletek, hipotézisek, mérések, adatok 2010, Budapest, Gondolat 135 o

Mészáros József: Bevezetés a társadalmi választások elméletébe 2008, Budapest, Gondolat 150 o

Könyvrészlet:

Mészáros József – Szakadát István: Ez elmúlt húsz év választásai in:Magyarország Politikai

Évhuszadkönyve 2009 DKMKA szerk. Sándor Péter, Stumpf Anna, Vass László

Kiadványrészlet:

Mészáros József :Az ökológiai kapcsolat problémája, Konferenciakötet GEO 2007

Cikk:

Mészáros József, Solymosi Norbert: The territorial distribution of Hungarian election data, Political Geography 2007, 26. 804-823

Mészáros József, Molnár D. László: Az aggregált adatok felbontásának lehetőségei (az ökológiai kapcsolat problémája) , Szociológiai Szemle 2006/3

Kutatási jelentések:

Mészáros József: A szociológia módszertanának sajátosságai Kézirat 16 o.

Mészáros József: Elméletek, hipotézisek adatok Kézirat 16 o.

Mészáros József: A reprezentativitás fogalmáról, Kézirat 14 o.

Mészáros József Simpson Paradoxon Kézirat 12 o.

Mészáros József – Solymosi Norbert: Bevezetés Bayes-i statisztikába, Kézirat 14 o

Molnár D. László: Térbeli statisztikai mintavétel, Kézirat 30 o.

Molnár D. László: Krigelés, Kézirat 30 o.

Molnár D. László: Idősorelemzés és területi statisztika, Kézirat 26 o.

Molnár D. László,Aggregált adatok dekomponálása, Kézirat 28 o.

Solymosi Norbert, Molnár D. László, Mészáros József: Területi adatok statisztikai elemzése, Kézirat 16 o.

Solymosi Norbert: Bayes-i statisztikai következtetés, Kézirat 22 o.